On Discontinuous Galerkin Methods

O.C. Zienkiewicz*
Department of Civil Engineering
University of Wales, Swansea, UK

R.L. Taylor†
Department of Civil and Environmental Engineering
University of California at Berkeley, USA
e-mail: rlt@ce.vulture.berkeley.edu

Abstract

Discontinuous Galerkin methods have received considerable attention in recent years for applications to many problems in which convection and diffusion terms are present. Several alternatives for treating the diffusion flux effects have been introduced, as well as, for treatment of the convective flux terms. This report summarizes some of the treatments that have been proposed. It also considers how elementary finite volume methods may be considered as the most primative form of a discontinuous Galerkin method as well as how it may be formed as a finite element method. Several numerical examples are included in the report which summarize results for discontinuous Galerkin solutions of one-dimensional problems with a scalar variable. Results are presented for diffusion-reaction problems, convection-diffusion problems, and a special problem with a turning point. We identify aspects which relate to accuracy as well as stability of the method.

*Unesco Professor, CIMNE, UPC, Barcelona, Spain.
†Visiting Professor, CIMNE, UPC, Barcelona, Spain.

1 Introduction

The work by B.G. Galerkin on approximate solution of differential equations appears in the literature for the first time in 1915 in a paper discussing series solutions of rods and plates.[1] Apparently, he was a civil engineer at the Petersburg Technological Institute. This work, which complemented the earlier work of Rayleigh and Ritz, was addressed to similar problems and at the time did not make a great stir. Some use of the method is made by S. Crandall in his book on Engineering Analysis.[2] There the procedure of weighted residuals is discussed and here distinction between the various choices of weighting and trial functions is made. In these, Crandall attributes the name of Galerkin to one procedure alone – that is, the one in which the weighting and trial functions are identical.

Much later it was recognized that the Galerkin procedures form a basis of most, if not all, finite element formulations for both linear and non-linear problems (*e.g.*, see Chapter 3 of Reference [3]). However, for some categories of application it became necessary to distinguish the procedures which followed Crandall's definition of Galerkin methods and those in which alternative weighting functions could perform better. An example here was a proper model for convection and, for this problem, one of the first mathematicians dealing with it by finite element methods, Professor Ron Mitchell (Dundee), remarked that Galerkin methods can be divided into two categories: those associated with the name of Bubnov[4] for which equal interpolation and weighting is used and all the others in which they are not. The latter are taken to be the basis of Petrov-Galerkin methods[5] and the origin of both names can be found in the book on variational methods by S.G. Mihklin.[6] We are uncertain about the exact definition for the two approaches, since in conversations by the first author with some people at St. Petersburg University it appears that the reverse order could be made in this nomenclature.

We should point out that in all the early works cited above (except Mitchell's) it was assumed that the trial functions employed in the solution satisfied *all* boundary conditions of the problem addressed (e.g., of both Dirichlet and Neumann type) and the solution was constructed by merely multiplying the differential equation by the individual weight functions and integrating over the domain.[2, 6] This implies that the functions used in the approximations must possess derivatives to the order of the differential equation. When the first use of integratation by parts to lower the order of the derivatives appearing and to include the Neumann boundary conditions

as part of the resulting variational type equation is unclear. However, this added step is now universally accepted as part of the Galerkin solution procedure. Indeed, the name of Galerkin now survives in both Bubnov and Petrov forms and in many usages of finite elements has been associated with various additional adjectives.

The Characteristic Galerkin method, for instance, describes a procedure in which the concept of characteristics and the integration along that direction is important.[7] Other names, such as Taylor-Galerkin[8, 9], Galerkin Least Square[10], and the present one, Discontinuous Galerkin appear frequently. Of course if all finite element methods are of Galerkin type there is an infinite scope as all procedures can be so described. We have recently heard of another name, that of *perturbed Galerkin*. It was made primarily because of his fame and wideness of the work this might be a name for some applications. But more about discontinuous Galerkin methods.

The name of discontinuous Galerkin appears to have started to be used in the early 1980's, and to the authors knowledge the name first appears in a paper by Delfour and Trochu in 1978.[11] An analysis for the scalar hyperbolic problem is presented by Johnson and Pitkäranta[12] and later in the book by Johnson for parabolic problems.[13] However, what is the concept and what does the methodology present? Viewed from the current ideas it is the opinion of the authors that it represents a method of linking separate domains in which finite element, series, or whatever other current procedures of solution is used for approximation.

It is well known that such linking can be accomplished by addition of further functions, Lagrangian multipliers, at contiguous interfaces of the various domains (which in themselves might be either a single element or multiple element size). The first such Lagrangian procedures in a finite element context have been used by Pian and his associates[14, 15], however, the essence of discontinuous Galerkin lies in elimination of the Lagrangian multiplier so that the total number of variables remains the same as that in the individual regions. How can such elimination be made? The most obvious method perhaps is that of elimnation of the Lagrange multipliers is by a direct substitution of the variables. Such substitution could be made either in the final approximating equations but perhaps better in the variational principle (or weak form) from which they are derived. The first use of such an elimination was provided by a paper of Kikuchi and Ando[16] in which the process is used to restore slope compatibility between the incompatible elements introduced by Bazeley *et al.* in 1966.[17] The paper was much criticized and perhaps

did not make the impact it should have. An almost identical process was used by Nitsche[18] in a mathematical sense and he discovered that direct substitution can lead to numerical problems such as the singularity of resulting equation system or their indefiniteness. To avoid this Nitsche added a further imposition of the constraint by a least square process, introducing of course another parameter which can be considered today as simply one of stabililzation. The ideas expressed at that time did not appear very much used in solution of practical problems although they have been directly applied to the process of pure convection by Johnson and Pitkäranta [Reed and Hill[19] and Lesaint and Raviart[20]]. This of course leads to the solution of the first order equations occuring in time and presents a possibility of yet another finite element approximation in time. The nature of impetus to the work is contained in the practical applications introduced various authors for problems in fluid dynamics and aerodynamics [*e.g.*, see Karniadakis *et al.* [21, 22, 23] and Cockburn [24, 25]] and a recently published dissertation by C.E. Baumann[26], who together with Oden published a series of important papers.[27, 28, 29, 30, 31] In all of these precisely the notions explained above are used but the procedure of Oden and Baumann to stabilize the equation appears to be one of simply changing the sign of the terms ensuring the interelement compatibility for diffusive flux terms, which seems to avoid the difficulties for all elements with an order higher than two, but not those for an element of order one.

We shall describe the various applications and some of the mathematics of linking in the following sections. However, at the outset we would like to outline some of our general views about the practicality of the process – even though on occasions it may prove to achieve computational results of standard finite elements.

1. Actual physical discontinuities are only well modelled if by luck or by previous adaptive analysis, the interface of the elements is placed exactly on such discontinuity. Thus, the modeling of fluid mechanics shocks in high speed flow is not an immediate advantage.

2. For hyperbolic problem, to some extent, an easing of the instability due to convective terms is achieved. The most important way in which the methodology does it however, is in not enforcing the convection conditions exactly at the outlet of exit boundary for the problem. This is well known to achieve the desired result even if a standard finite element process is known. Thus, for instance, in the one-dimensional

problem all the oscillations disappear if the exit conditions are removed or simply the Neumann condition is imposed on the diffusive part.

3. The cost of using discontinuous Galerkin appears to be very large. In one-dimensional problems at the junction of any two subdomains the variables are doubled in number and this increase becomes much larger in two- and three-dimensions depending on the nature of the problem. The very description of the problem in such circumstances may be difficult though it appears to be achieved by quite well by some people. However, the perhaps future progress lies in a more limited application of discontinuous Galerkin only to areas where such discontinuities may occur.

2 Diffusion-Convection-Reaction Equations with Scalar field

We begin by considering the solution of the diffusion-convection-reaction equation expressed in terms of the *scalar* function u.

$$\frac{\partial u}{\partial t} - \boldsymbol{\nabla} \cdot [\boldsymbol{k}(\boldsymbol{x}) \, \boldsymbol{\nabla} \, u - \boldsymbol{a} \, u] + c(\boldsymbol{x}) \, u = q(\boldsymbol{x}) \tag{1}$$

Here $\boldsymbol{\nabla}$ denotes the gradient, $\boldsymbol{k}(\boldsymbol{x})$ is a symmetric matrix of diffusion coefficients, $\boldsymbol{a}(\boldsymbol{x})$ is a convection velocity, $c(\boldsymbol{x})$ is a reaction function, $q(\boldsymbol{x})$ a specified loading function and $() \cdot ()$ denotes the inner (dot) product between two vectors. The differential equation is assumed to be valid for all $\boldsymbol{x}$ in a domain Ω.

We define *fluxes* for diffusion and convection by

$$\boldsymbol{F}_d = \boldsymbol{k} \, \boldsymbol{\nabla} \, u \quad \text{and} \quad \boldsymbol{F}_c = \boldsymbol{a} \, u \ , \tag{2}$$

respectively. In the sequel the fluxes play an important roll in the construction of interface approximations between two contiguous subdomains (elements).

For the transient problem it is necessary to specify an initial condition. This may be written as

$$u(\boldsymbol{x}, 0) = u_0(\boldsymbol{x}) \quad ; \ \boldsymbol{x} \in \Omega \tag{3}$$

In addition to the initial condition boundary conditions are considered in two forms. These are

$$u(\boldsymbol{x}, t) = \bar{u}_d \ ; \ \boldsymbol{x} \in \Gamma_d \tag{4}$$

known as the *Dirichlet* condition, and

$$\boldsymbol{n} \cdot \boldsymbol{F}_d = \bar{F}_n \ ; \ \boldsymbol{x} \in \Gamma_n \tag{5}$$

known as the *Neumann* condition.

In the first part of this report we consider the steady-state problem given by

$$-\boldsymbol{\nabla} \cdot [\boldsymbol{k}(\boldsymbol{x}) \, \boldsymbol{\nabla} \, u - \boldsymbol{a} \, u] + c(\boldsymbol{x}) \, u = q(\boldsymbol{x}) \tag{6}$$

which obviously does not necessitate use of an initial condition.

2.1 Galerkin Solution - Weak Forms

To construct an approximate solution to the steady state problem given above we introduce a standard Galerkin procedure in which the differential equation is multiplied by an arbitrary weight function $\delta u(\boldsymbol{x})$. Subsequently, the diffusion term is integrated by parts to obtain the weak form:

$$B(\delta u, u) = L(\delta u) \tag{7}$$

where

$$\begin{aligned} B(\delta u, u) \ = \ & \int_\Omega \left\{ (\boldsymbol{\nabla} \delta u)^T \, [\boldsymbol{k} \boldsymbol{\nabla} u] + \delta u \, [\boldsymbol{\nabla} \cdot (\boldsymbol{a} u) + c \, u] \right\} \mathrm{d}\Omega \\ & - \int_\Gamma \delta u \ [\boldsymbol{n} \cdot \boldsymbol{F}_d] \ \mathrm{d}\Gamma \end{aligned} \tag{8}$$

and

$$L(\delta u) = \int_\Omega \delta u \, q \, \mathrm{d}\Omega \tag{9}$$

For this form to be valid the weight function δu and any approximation to the solution variable u must be at least C^0 continuous in Ω, which is some region of interest and may be one or more elements in a finite element representation.

Accordingly, here we first introduce the standard Galerkin approximation

$$u(\boldsymbol{x}) = \mathbf{N}(\boldsymbol{x})\,\tilde{\mathbf{u}} \quad \text{and} \quad \delta u(\boldsymbol{x}) = \mathbf{N}(\boldsymbol{x})\,\delta\tilde{\mathbf{u}} \tag{10}$$

which we shall assume are p-order polynomials in Ω. Evaluation of integrals appearing in Eqs (8) and (9) gives the matrix problem:

$$\mathbf{H}\tilde{\mathbf{u}} - \int_{\Gamma} \mathbf{N}^T F_n \,\mathrm{d}\Gamma = \mathbf{f} \tag{11}$$

where

$$\mathbf{H} = \int_{\Omega} \underbrace{(\boldsymbol{\nabla}\mathbf{N})^T\, \boldsymbol{k}\boldsymbol{\nabla}\mathbf{N}}_{Diffusion}\,\mathrm{d}\Omega + \int_{\Omega} \underbrace{\mathbf{N}^T c\mathbf{N}}_{Reaction}\,\mathrm{d}\Omega + \int_{\Omega} \underbrace{(\mathbf{N})^T\,[\boldsymbol{\nabla}\cdot(\boldsymbol{a}\mathbf{N})]}_{Convection}\,\mathrm{d}\Omega \tag{12}$$

and

$$\mathbf{f} = \int_{\Omega} \mathbf{N}^T q \,\mathrm{d}\Omega \,. \tag{13}$$

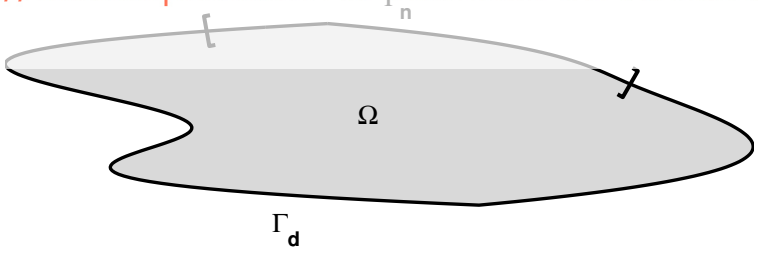

Figure 1: Single domain and its boundary parts

For *single* domain problems we divide the boundary into the two parts

$$\Gamma_d \cup \Gamma_n = \Gamma \quad ; \quad \Gamma_d \cap \Gamma_n = 0$$

as shown in Fig. 1 Accordingly, we split the boundary integral in Eq. (12) and introduce the known value for the Neumann part of the boundary to

obtain

$$\int_{\Gamma} \mathbf{N}^T F_n \, d\Gamma = \underbrace{\int_{\Gamma_d} \mathbf{N}^T F_n \, d\Gamma}_{Omit:\ \delta u=0} + \int_{\Gamma_n} \mathbf{N}^T \bar{F}_n \, d\Gamma \ . \tag{14}$$

The matrix problem for a single domain may then be written as

$$\mathbf{H}\tilde{\mathbf{u}} = \mathbf{f} + \int_{\Gamma_n} \mathbf{N}^T \bar{F}_n \, d\Gamma = \tilde{\mathbf{f}} \tag{15}$$

to which we also must impose $u = \bar{u}$ on Γ_d.

2.2 Multiple Domains

Figure 2: Two subdomains with an interface boundary

Consider next a problem which is divided into multiple subdomains. It is sufficient to consider the *two* subdomain problem shown in Fig. 2. Here there are two sets of Galerkin equations which we denote as:

$$\begin{aligned} \mathbf{H}^1\tilde{\mathbf{u}}^1 - \textstyle\int_\Gamma \mathbf{N}_1^T F_n^1 \, d\Gamma &= \mathbf{f}^1 \\ \mathbf{H}^2\tilde{\mathbf{u}}^2 - \textstyle\int_\Gamma \mathbf{N}_2^T F_n^2 \, d\Gamma &= \mathbf{f}^2 \end{aligned} \tag{16}$$

We now divide each subdomain boundary into *three* parts

$$\Gamma_d^i \cup \Gamma_n^i \cup \Gamma_{int} = \Gamma^i \quad ; \quad i = 1, 2 \tag{17}$$

in which Γ_{int} is the interface boundary of contiguous subdomains (Figure 3).

In any solution procedure we must satisfy, at least approximately, a condition on continuity for u which may be expressed by

$$u^1 = u^2 \tag{18}$$

as well as a condition on flux equilibrium given by

$$\boldsymbol{n}^1 \cdot [\boldsymbol{k}^1 \boldsymbol{\nabla} u^1] + \boldsymbol{n}^2 \cdot [\boldsymbol{k}^2 \boldsymbol{\nabla} u^2] = 0 \tag{19}$$

and this must be enforced along each interface Γ_{int}.

2.3 Lagrange Multiplier Solution

As a first procedure to enforce the continuity and flux equilibrium along each interface we consider a classical lagrangian multiplier method. Accordingly, we introduce the Lagrange multiplier λ on Γ_{int} as

$$\lambda = \boldsymbol{n}^1 \cdot [\boldsymbol{k}^1 \nabla u^1] = -\boldsymbol{n}^2 \cdot [\boldsymbol{k}^2 \nabla u^2] \tag{20}$$

Clearly, this automatically satisfies the flux equilibrium equation given by Eq. (19). If, further, we add to $B(\delta u, u)$ a classical multiplier form

$$B_\lambda(\delta u, \delta\lambda, u, \lambda) = \int_{\Gamma_{int}} [\delta u^2 - \delta u^1]\, \lambda \, d\Gamma + \int_{\Gamma_{int}} \delta\lambda\, [u^2 - u^1]\, d\Gamma \tag{21}$$

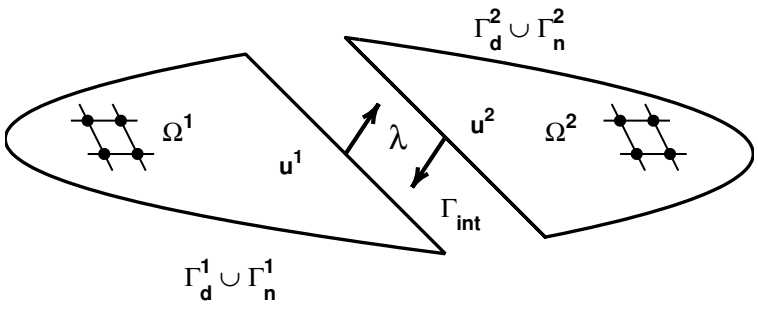

Figure 3: Two subdomains with Lagrange multiplier on interface boundary

together with the approximation for the multiplier as

$$\lambda = \mathbf{N}_\lambda \tilde{\boldsymbol{\lambda}} \tag{22}$$

we obtain from Eq. (16) the matrix problem

$$\begin{aligned} \mathbf{H}^1 \tilde{\mathbf{u}}^1 - \mathbf{C}^1 \tilde{\boldsymbol{\lambda}} &= \tilde{\mathbf{f}}^1 \\ \mathbf{H}^2 \tilde{\mathbf{u}}^2 + \mathbf{C}^2 \tilde{\boldsymbol{\lambda}} &= \tilde{\mathbf{f}}^2 \end{aligned} \tag{23}$$

where as before $\tilde{\mathbf{f}}$ now includes the Neumann boundary term,

$$\mathbf{C}^1 = \int_{\Gamma_{int}} \mathbf{N}_1^T \mathbf{N}_\lambda \, \mathrm{d}\Gamma \text{ and } \mathbf{C}^2 = \int_{\Gamma_{int}} \mathbf{N}_2^T \mathbf{N}_\lambda \, \mathrm{d}\Gamma \ . \tag{24}$$

To these we add the continuity condition as:

$$\int_{\Gamma_{int}} \delta\lambda \, \left[u^2 - u^1 \right] \mathrm{d}\Gamma = 0 \tag{25}$$

Substituting the approximations for the fields we get the added matrix set of constraints

$$-(\mathbf{C}^1)^T \tilde{\mathbf{u}}^1 + (\mathbf{C}^2)^T \tilde{\mathbf{u}}^2 = \mathbf{0} \tag{26}$$

The above steps may be written compactly as:

$$\begin{bmatrix} \mathbf{H}^1 & \mathbf{0} & -\mathbf{C}^1 \\ \mathbf{0} & \mathbf{H}^2 & \mathbf{C}^2 \\ -(\mathbf{C}^1)^T & (\mathbf{C}^2)^T & \mathbf{0} \end{bmatrix} \begin{Bmatrix} \tilde{\mathbf{u}}^1 \\ \tilde{\mathbf{u}}^2 \\ \tilde{\boldsymbol{\lambda}} \end{Bmatrix} = \begin{Bmatrix} \tilde{\mathbf{f}}^1 \\ \tilde{\mathbf{f}}^2 \\ \mathbf{0} \end{Bmatrix} \tag{27}$$

in which $\tilde{\boldsymbol{\lambda}}$ is the extra set of variables from the lagrangian multipliers. This set is a standard *mixed form* and the coefficient matrix is clearly *indefinite*.

Any solution to the problem must enforce conditions for the mixed problem and these include a count condition given by[3]

$$n_1 + n_2 \geq n_\lambda \tag{28}$$

in which n_i denotes the number of free parameters in each u^i set. In addition, the equations must always have a unique solution. For further details on these requirements see Reference [3].

In the sequel we seek methods to replace $\tilde{\boldsymbol{\lambda}}$ by some expression in the remaining variables, $\tilde{\mathbf{u}}^i$. This is the basis of all *domain decomposition* methods (e.g., DtN, overlapping domains, mortaring, etc.). It is also the goal of *discontinuous Galerkin* methods.

2.4 Discontinuous Galerkin Methods

Numerous methods have been proposed by proponents of the discontinuous Galerkin method for expressions which replace λ by some expression in $\boldsymbol{u}^1$ and/or $\boldsymbol{u}^2$. Here we consider only some of these together with an evaluation of results attained from each.

2.4.1 Solutions for diffusion term

For *diffusion terms*, one of the simplest is to replace λ by the average of the flux from the contiguous elements along the interface boundary Γ_{int}. Accordingly, we can write

$$\lambda = \frac{1}{2}\,\mathbf{n}\cdot\left[\boldsymbol{k}^1\boldsymbol{\nabla}u^1 + \boldsymbol{k}^2\boldsymbol{\nabla}u^2\right] \tag{29}$$

in which $\mathbf{n}$ is an outward normal to one of the domains. Introducing this into the Galerkin equations for each domain gives first two sets of equations as

$$\mathbf{H}^1\tilde{\mathbf{u}}^1 \;-\; \mathbf{C}^{11}\tilde{\mathbf{u}}^1 - \mathbf{C}^{12}\tilde{\mathbf{u}}^2 = \mathbf{f}^1 \tag{30}$$

$$\mathbf{H}^2\tilde{\mathbf{u}}^2 + \mathbf{C}^{21}\tilde{\mathbf{u}}^1 + \mathbf{C}^{22}\tilde{\mathbf{u}}^2 = \mathbf{f}^2 \tag{31}$$

At this stage the method is consistent, but usually singular or very ill-conditioned. It is necessary to include the effects from the constraint equation

$$(\delta\tilde{\boldsymbol{\lambda}})^T\left[-\mathbf{C}^1\tilde{\mathbf{u}}^1 + \mathbf{C}^1\tilde{\mathbf{u}}^1\right] = 0\ . \tag{32}$$

which is the matrix expression resulting from the condition Eq. (25). This condition immediately suggests using a similar condition to replace $\delta\lambda$ by a variation on the average flux. The final result of such substitution is the pair of equations

$$\begin{aligned}\left[\mathbf{H}^1 - \mathbf{C}^{11} - \mathbf{C}^{11,T}\right]\tilde{\mathbf{u}}^1 + \left[\mathbf{C}^{21,T} - \mathbf{C}^{12}\right]\tilde{\mathbf{u}}^2 &= \tilde{\mathbf{f}}^1\\ \left[\mathbf{C}^{21} - \mathbf{C}^{12,T}\right]\tilde{\mathbf{u}}^1 + \left[\mathbf{H}^2 + \mathbf{C}^{22} + \mathbf{C}^{22,T}\right]\tilde{\mathbf{u}}^2 &= \tilde{\mathbf{f}}^2\end{aligned} \tag{33}$$

or

$$\left[\boldsymbol{H} + \mathbf{C} + \mathbf{C}^T\right]\ \tilde{\mathbf{U}} = \tilde{\mathbf{F}}\ ;\quad \tilde{\mathbf{U}} = (\tilde{\boldsymbol{u}}^1\ ,\ \tilde{\boldsymbol{u}}^2)^T$$

Returning to the original Galerkin form taken from Eq. (12) and extended to two subdomains Ω_1 and Ω_2 indicates the steps accomplished in the replacement of the multipliers for the diffusion treatment.

We define an average flux using the notation

$$\left\langle \mathbf{k}\boldsymbol{\nabla} u \right\rangle = \frac{1}{2}\left(\boldsymbol{k}^1\boldsymbol{\nabla} u^1 + \boldsymbol{k}^2\boldsymbol{\nabla} u^2\right) \tag{34}$$

and a *jump* in the solution at the interface by [1]

$$[\![u\,\boldsymbol{n}]\!] = u^1\boldsymbol{n}^1 + u^2\boldsymbol{n}^2 \tag{35}$$

with similar expressions for the variation.

For a two subdomain problem with a single interface boundary Γ_{int}. the Galerkin equations may be written as (where $\boldsymbol{a}$ is assumed for now as zero):

$$\begin{aligned} B(\delta u, u) \;=\; & \sum_{i=1}^{2}\int_{\Omega^i}\Big\{\underbrace{\left(\boldsymbol{\nabla}\delta u^i\right)^T\left[\boldsymbol{k}^i\boldsymbol{\nabla} u^i\right] + \delta u^i\left[c\,u^i\right]}_{Two\ bodies}\Big\}\,\mathrm{d}\Omega \\ & - \int_{\Gamma_{int}}\underbrace{[\![\delta u\,\boldsymbol{n}]\!]\cdot\left\langle \boldsymbol{k}\boldsymbol{\nabla} u\right\rangle}_{Flux\ balance}\,\mathrm{d}\Gamma + \alpha\int_{\Gamma_{int}}\underbrace{\left\langle \boldsymbol{k}\boldsymbol{\nabla}\delta u\right\rangle\cdot[\![u\,\boldsymbol{n}]\!]}_{Continuity\ of\ u}\,\mathrm{d}\Gamma \end{aligned} \tag{36}$$

(37)

and

$$L(\delta u) = \sum_{i=1}^{2}\int_{\Omega^i}\delta u^i\, q^i\,\mathrm{d}\Omega \tag{38}$$

For standard replacement of Lagrange multipliers we set $\alpha = -1$, however, it is most important to note that α can be other choices.

Nitsche used $\alpha = -1$ to impose Dirichlet boundary conditions as natural condions in a variational problem.[18] In this effort it was necessary to include an added least-square term which *stabilized* the result. In the context of a discontinuous interface condition the resulting expressions are as above but with the added term introduced into the weak form as

$$\int_{\Gamma_{int}} \tau\;[\![\delta u\,\boldsymbol{n}\,]\!]\cdot[\![\,u\,\boldsymbol{n}\,]\!]\;\mathrm{d}\Gamma$$

[1] Different methods are given in the literature for defining the flux and jump separation, however, the above appears to be particularly useful when coding the discontinuous Galerkin method.

where τ is the stabilization parameter which Nitsche finds must be $O(|k|/h) > 0$ in which h is an element size measure and $|k|$ is a norm of the diffusion matrix.

As an alternative Oden & Baumann[27, 29, 30] choose:

$$\alpha = 1 \ .$$

This method leads to the modified form

$$\left[\boldsymbol{H} + \mathbf{C} - \mathbf{C}^T\right] \tilde{\mathbf{U}} = \mathbf{F} \tag{39}$$

and it is found the coefficient matrix is often stable for $\tau = 0$ when polynomial orders of two or more are used to approximate the u^i in each subdomain.

To understand why $\alpha = 1$ is a more stable approximation consider a one-dimenaional problem where we will obtain a form

$$\mathbf{C} - \mathbf{C}^T \propto \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

which obviously is a term with a *positive* determinant. Generally, in multiple dimensions Oden & Baumann find that the interface terms are stable with $\tau = 0$.

2.4.2 Treatment of convection terms

We now return to the case where $\boldsymbol{a}$ is not zero. In this case the convection term in the Galerkin equation (8) is given by

$$B_c(\delta u, u) = \int_\Omega \delta u \, \boldsymbol{\nabla} \cdot [\boldsymbol{a} u] \, \mathrm{d}\Omega \tag{40}$$

where the convective flux is expressed by $\boldsymbol{F}_c = \boldsymbol{a} u$. Normally, in approximate solutions this term is not integrated by parts, however, in developing a treatment within the context of a discontinuous Galerkin process we do integrate it by parts to obtain

$$B_c(\delta u, u) = \int_\Omega \delta u \, \boldsymbol{\nabla} \cdot [\boldsymbol{a} u] \, \mathrm{d}\Omega = -\int_\Omega (\boldsymbol{\nabla} \delta u)^T \, \boldsymbol{a} u \, \mathrm{d}\Omega + \int_\Gamma \delta u \, \boldsymbol{n} \cdot \boldsymbol{F}_c \, \mathrm{d}\Gamma \tag{41}$$

In this form a discontinuous Galerkin process can treat the convective flux by a process similar to that used for $\boldsymbol{F}_d$.

Thus, we again consider two domains where the terms become

$$\begin{aligned} B_c(\delta u, u) &= \sum_{i=1}^{2} \int_{\Omega^i} \delta u^i \, \boldsymbol{\nabla}^T \left(\boldsymbol{a}^i u^i \right) \, \mathrm{d}\Omega \\ &= -\sum_{i=1}^{2} \int_{\Omega^i} (\boldsymbol{\nabla} \delta u^i)^T \, \boldsymbol{a}^i u^i \, \mathrm{d}\Omega + \int_{\Gamma} [\![\delta u \boldsymbol{n}]\!]^T \left\langle \boldsymbol{F}_c \right\rangle \mathrm{d}\Gamma \end{aligned} \tag{42}$$

in which $[\![\delta u \, \boldsymbol{n}]\!] = \delta u^1 \, \boldsymbol{n}^1 + \delta u^2 \, \boldsymbol{n}^2$.

Treatment of the boundary term now involves the solution field u directly, hence it is necessary only to find a consistent replacement for the convective flux. A good approximation for $\left\langle \boldsymbol{F}_c \right\rangle$ on each interface is the *outflow* value from a domain. Thus, all boundaries of subdomains are examined according to the behavior of the normal flux and separated into the two categories:

$$\begin{aligned} \boldsymbol{n} \cdot \boldsymbol{a} &> 0 \Rightarrow \text{outflow} \\ \boldsymbol{n} \cdot \boldsymbol{a} &< 0 \Rightarrow \text{inflow} \end{aligned}$$

The value of the convective flux on the outflow boundary [denoted as the minus $(-)$ boundary] is then used for the approximation

$$\left\langle \boldsymbol{F}_c \right\rangle = \boldsymbol{a}^- u^- \tag{43}$$

The process is shown conceptually for a one-dimensional problem in Fig. (4).

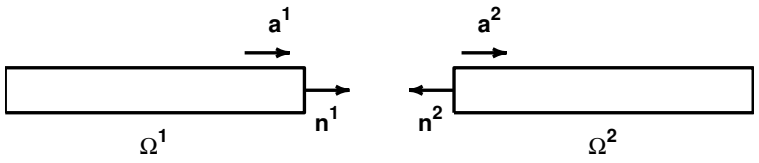

Figure 4: Treatment of convective flux in one-dimension

Often for implementation the volume term in Eq. (42) is integrated by

parts again to get the alternative expression convective term as

$$B_c(\delta u, u) = \sum_{i=1}^{2} \int_{\Omega^i} \delta u^i \, \boldsymbol{\nabla}^T \left(\boldsymbol{a}^i u^i \right) \, \mathrm{d}\Omega + \int_{\Gamma_{int}} \delta u^+ \left[\boldsymbol{n}^+ \cdot (\boldsymbol{a}^- u^- - \boldsymbol{a}^+ u^+) \right] \, \mathrm{d}\Gamma \tag{44}$$

In this form terms on boundaries appear only on inflow and interface parts. On an inflow boundary point we use the Neumann boundary condition and set: $\boldsymbol{n}^+ \cdot [\boldsymbol{a}^+ u^+] \rightarrow \bar{F}_n$. Inflow boundary terms then are added to the $L(\delta u)$ given in Eq. (9).

Treatment of the convective terms by the above process amounts to an *upwind* treatment at the interfaces. We shall observe in the examples, however, that no such upwind treatment is available within each subdomain. Thus, within each subdomain additional treatment may be needed to control spurrious oscillations.[7]

2.5 Discontinuous Galerkin weak form

Collecting all the above terms together we obtain the final form for the solution of a diffusion-convection-reaction equation by a discontinuous Galerkin procedure. The result is

$$\begin{aligned} B(\delta u, u) \;=\; & \sum_{i=1}^{2} \int_{\Omega^i} \left\{ \left(\boldsymbol{\nabla} \delta u^i \right)^T \left[\boldsymbol{k}^i \boldsymbol{\nabla} u^i \right] + \delta u^i \left[c \, u^i \right] + \delta u^i \, \boldsymbol{\nabla}^T \left(\boldsymbol{a}^i u^i \right) \right\} \mathrm{d}\Omega \\ - \; & \int_{\Gamma_{int}} [\![\delta u \, \boldsymbol{n}]\!] \cdot \left\langle \boldsymbol{k} \boldsymbol{\nabla} u \right\rangle \mathrm{d}\Gamma + \alpha \int_{\Gamma_{int}} \left\langle \boldsymbol{k} \boldsymbol{\nabla} \delta u \right\rangle \cdot [\![u \, \boldsymbol{n}]\!] \, \mathrm{d}\Gamma \\ + \; & \int_{\Gamma_{int}} \delta u^+ \left[\boldsymbol{n}^+ \cdot (\boldsymbol{a}^- u^- - \boldsymbol{a}^+ u^+) \right] \, \mathrm{d}\Gamma \end{aligned} \tag{45}$$

and

$$L(\delta u) = \sum_{i=1}^{2} \left\{ \int_{\Omega^i} \delta u^i \, q^i \, \mathrm{d}\Omega + \int_{\Gamma_n} \delta u^i \, \bar{F}_n^i \mathrm{d}\Gamma \right\} \tag{46}$$

2.6 One-dimensional Diffusion-Advection-Reaction Problems

The differential equation for a scalar second order diffusion-advection-reaction equation in one dimension may be written as

$$-\frac{d}{dx}\left(k(x)\,\frac{du}{dx}\right)+\frac{d}{dx}\Big(a(x)\,u\Big)+c(x)\,u=q(x)\quad;\quad x\in\Omega \tag{47}$$

where $k(x)$ is the diffusion coefficient, $a(x)$ is an advection coefficient, $c(x)$ is a reaction coefficient, u is the scalar unknown variable, and q is a given loading function.

Boundary conditions may be given as specified u (Dirichlet type)

$$u(x)=\bar{u}\quad;\quad x\in\Gamma_d \tag{48}$$

or specified flux (Neumann type)

$$F_d(x)=k\,\frac{du}{dx}=\bar{F}_n\quad;\quad x\in\Gamma_n \tag{49}$$

where $\Gamma=\Gamma_d\cup\Gamma_n$ and $\Gamma_d\cap\Gamma_n=0$ with Γ the total boundary of domain Ω.

2.6.1 Weak form in one-dimension

The terms in the weak form for a discontinuous Galerkin solution of the problem given in Eqs. (47) to (49) may be written as

$$\begin{aligned}
B(u,w) &= \int_{\Omega_e}\left[\frac{dw}{dx}\,k\,\frac{du}{dx}-\frac{dw}{dx}\,a\,u+w\,c\,u\right]\mathrm{d}x\\
&- \Big[w\,F_d(u)\Big]_{\Gamma_d}+\alpha\Big[w\,F_d(u)\Big]_{\Gamma_d}\\
&- \Big[\,[\![w]\!]\,\langle F_d(u)\rangle\Big]_{\Gamma_{int}}+\alpha\Big[\,[\![u]\!]\,\langle F_d(w)\rangle\Big]_{\Gamma_{int}}\\
&+ \Big[w\,a^-\,u\Big]_{\Gamma^-}+\Big[w\,a^-\,u\Big]_{\Gamma^-_{int}}\\
&+ \tau\Big[w\,u\Big]_{\Gamma_d}+\alpha\Big[F_d(w)\,\bar{u}\Big]_{\Gamma_d}+\Big[\,[\![w]\!]\,[\![u]\!]\,\Big]_{\Gamma_{int}}
\end{aligned} \tag{50}$$

and

$$\begin{aligned}
L(w) &= \int_{\Omega_e} w\,f\,\mathrm{d}x+\alpha\Big[F_d(w)\,\bar{u}\Big]_{\Gamma_d}+\Big[w\,\bar{F}_d\Big]_{\Gamma_n}\\
&- \Big[w\,a^-\,u\Big]_{\Gamma^-}+\tau\Big[w\,\bar{u}\Big]_{\Gamma_d}
\end{aligned} \tag{51}$$

In the above Ω_e denotes the integral over the interior of elements and Γ_{int} is the interface between contiguous elements. The terms within the special brackets are interpretted as the jump

$$\llbracket u \rrbracket = u^+ n^+ + u^- n^- \tag{52}$$

where u^+ is the value on the right + boundary; u^- the value on the $-$ side boundary; n^+ is the outward pointing normal to the + boundary and n^- is the outward pointing normal on the $-$ boundary. Similarly,

$$\langle F \rangle = \frac{1}{2} \left(F^+ + F^- \right) \tag{53}$$

where F^+ is the flux from the + element and F^- the flux from the $-$ element. Satisfaction of these two conditions exactly implies continuity of the solution u and balance of the flux F at the interfaces.

The parameter α takes the values -1 or 1 depending on the particular form of the discontinuous Galerkin scheme to be used. Use of $\alpha = -1$ gives a fully symmetric tangent for the diffusion term, whereas use of $\alpha = 1$ gives the form used by Oden and Baumann.[31] Finally, the parameter τ is used to provide better stability to the approximation.[32]

3 Transient Problems

Until now we have not considered the solution of transient problems, however, it is in this class of problems that a very desirable feature of the discontinuous Galerkin method is obtained.

For the second order problem considered, a transient first order time derivative appears in the Galerkin solution as

$$\frac{\partial u}{\partial t} \quad \rightarrow \quad \sum_{i=1}^{2} \int_{\Omega^i} \delta u \left[\frac{\partial u}{\partial t} \right] \mathrm{d}\Omega \tag{54}$$

This is shown for the two-body case but as can observe generalizes for any number of bodies merely by extending the sum over the number of subregions (elements) used.

Introducing a standard finite element approximation written in the separable form

$$u(\boldsymbol{x}, t) = \boldsymbol{N}(\boldsymbol{x})\, \tilde{\boldsymbol{u}}(t) \tag{55}$$

the time derivatives may be computed by differentiating the parameters $\tilde{\boldsymbol{u}}$. Evaluation of the integrals in space gives the semi-discrete form of the first derivative term as

$$\sum_{i=1}^{2} \int_{\Omega^i} \delta u \left[\frac{\partial u}{\partial t}\right] \mathrm{d}\Omega = \left[\begin{array}{cc} \delta\tilde{\boldsymbol{u}}^1 & \delta\tilde{\boldsymbol{u}}^2 \end{array}\right] \left[\begin{array}{cc} \mathbf{M}^{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{M}^{22} \end{array}\right] \left\{\begin{array}{c} \frac{\partial\tilde{\boldsymbol{u}}^1}{\partial t} \\ \frac{\partial\tilde{\boldsymbol{u}}^2}{\partial t} \end{array}\right\} \tag{56}$$

We immediately observe that there is *no coupling* between the rate terms in the two subregions. This generalizes for an N-body problem to

$$\sum_{i=1}^{N} \int_{\Omega^i} \delta u \left[\frac{\partial u}{\partial t}\right] \mathrm{d}\Omega = \left[\begin{array}{cccc} \delta\tilde{\boldsymbol{u}}^1 & \delta\tilde{\boldsymbol{u}}^2 & \ldots & \delta\tilde{\boldsymbol{u}}^N \end{array}\right] \left[\begin{array}{cccc} \mathbf{M}^{11} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{M}^{22} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{M}^{NN} \end{array}\right] \left\{\begin{array}{c} \delta\tilde{\boldsymbol{u}}^1 \\ \delta\tilde{\boldsymbol{u}}^2 \\ \ldots \\ \delta\tilde{\boldsymbol{u}}^N \end{array}\right\} \tag{57}$$

We note that the form is particularly useful for any *explicit* solution method as any p-order approximation leads to a block diagonal coefficient to all rate terms. Indeed, for the structure above, the matrix may be made completely *diagonal* by using shape functions which are orthogonal on each element. For a class of elements such orthogonal shape functions have been deduced by Karniadakis *et al.*[21, 22, 33]

It is also possible to use discontinuous Galerkin methods in *time*. The process follows precisely the approach for convective terms and has been exploited quite early by Johnson[34, 13, 3] However, most applications to date which use the discontinuous Galerkin method for spatial discretization use standard integrations methods[3] or those of the Runge-Kutta type.[33]

4 Finite Volume Methods and their Relation to Discontinuous Galerkin Procedures

It is the view of some users that the discontinuous Galerkin process has much in common with the finite volume process. In particular, as both satisfy what is apparently known as *local conservativity* on each domain (element) individually.[33] It is perfectly true that if the integral of inflows and outflows is taken over a finite volume (or over a finite element derived by the discontinuous Galerkin process) it will be found that exact satisfaction

of local conservativity conditions is achieved. But that does not mean the fluxes recorded are correct or even accurate. It simply is a record of the fact that what goes into the artificial element or cell is precisely balanced by what goes out and/or is is stored. The authors do not view the fact that there is any merit in such local conservativity if the alternative procedure of finite elements results in more accurate fluxes at every stage of the calculation. Clearly such fluxes have to be evaluated by suitable recovery from the gradients which devise an element but this is suitably described in texts[3] and can always be shown to be true.

4.1 Cell centered method

In this section we would like to indicate that the finite volumes as frequently used are simply discontinuous Galerkin elements wherein very low order approximations are assumed to represent the variable in the cell. As the linear discontinuous element of one-dimensional or two-dimensional types has already been discussed and found to be quite sophisticated and accurate, the only possibility here is to look at zero order expansions with an element.

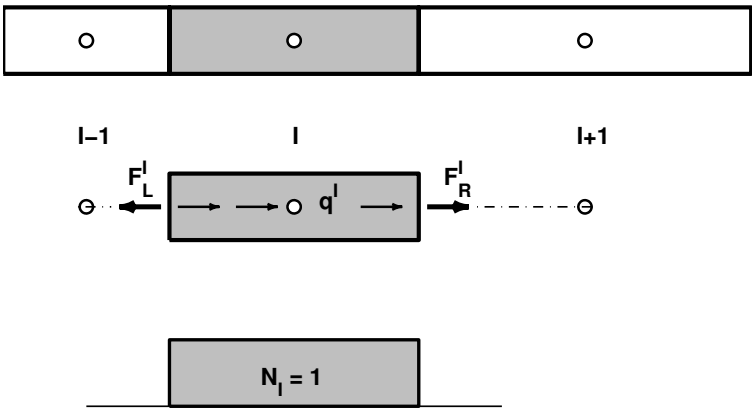

Figure 5: Cell centered finite volume in one dimension

In figure 5 we consider several cells of linear type in which a node I is indicated at the center simply to record the value of the function u within

each element and this is presumed to be constant there. We are thus talking about shape functions which are simply unity and indeed we shall use weight functions of the same kind.

In the figure we indicate very unequal size elements show fluxes where elements or cells are joined. Considering now cell I, shown in Fig. 5 as shaded, after integration by parts of the Galerkin form for the equation

$$\frac{\mathrm{d}}{\mathrm{d}x}\left[k\,\frac{\mathrm{d}u}{\mathrm{d}x}+a\,u\right]=q(x) \tag{58}$$

and evaluating the resulting integrals for constant approximation we obtrain (see Fig. 5)

$$1\cdot F_R-1\cdot F_L=\bar{q}\,h_I \tag{59}$$

where the values of unity arise from the constant shape function and F_L and F_R are total fluxes arising form convection and diffusion which are given by

$$F_R=\left[\underbrace{k\,\frac{\mathrm{d}u}{\mathrm{d}x}}_{Diffusion}+\underbrace{a\,u}_{Convection}\right]_R \tag{60}$$

At this stage it is of course impossible to evaluate the diffusive fluxes at the ends of cell I but a simple finite difference approximation will give the average gradients as

$$\left.\frac{\mathrm{d}u}{\mathrm{d}x}\right|_R=\frac{\tilde{u}^{I+1}-\tilde{u}^I}{(h_{I+1}+h_I)/2} \tag{61}$$

and

$$\left.\frac{\mathrm{d}u}{\mathrm{d}x}\right|_L=\frac{\tilde{u}^{I}-\tilde{u}^{I-1}}{(h_{I}+h_{I-1})/2} \tag{62}$$

In the same way, for the convective part we use the *upwind* value as in previous discontinuous Galerkin approximation. Assuming k and a are positive and constant we obtain the approximation for the total flux F_R as

$$F_R=k\,\frac{\tilde{u}^{I+1}-\tilde{u}^I}{(h_{I+1}+h_I)/2}+a\,\tilde{u}^I \tag{63}$$

and for F_L the approximation

$$F_L = k\,\frac{\tilde{u}^I - \tilde{u}^{I-1}}{(h_I + h_{I-1})/2} + a\,\tilde{u}^{I-1} \tag{64}$$

Substitution into Eq. (59) gives

$$k\,\left[\frac{\tilde{u}^{I+1} - \tilde{u}^I}{(h_{I+1} + h_I)/2} - \frac{\tilde{u}^I - \tilde{u}^{I-1}}{(h_I + h_{I-1})/2}\right] + a\,\left[\tilde{u}^I - \tilde{u}^{I-1}\right] = \bar{q}\,h_I \tag{65}$$

and it is obvious that *full upwinding* of the convection part has occurred and the diffusion part replaced by the average central difference approximation on unequal intervals.

The same procedures can be extended to two and three-dimensional elements and in Fig. 6 we show two triangles and their interface. The approximation are of course a little more difficult.

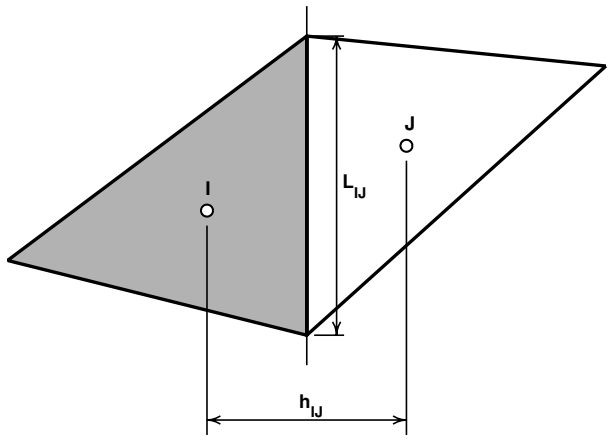

Figure 6: Cell centered finite volume in two dimension

The cell centered finite volume method in two dimensions requires a balance for each subdomain in which, again, the approximations are assumed to be constant. All procedes as in the one dimensional case except the area of the subdomain replaces the h_I multiplying the $\bar{q}$ and flux balance is carried out over sides of the subdomain. For example, using simple triangles, as shown in Fig. 6, it is necessary to approximate the integral of the flux along

the contiguous interface. For simplicity we have shown this as a vertical side of length L_{IJ} in the figure.

Writing the flux balance we have for the I element the contribution for the flux along the IJ-side as

$$F_{IJ}\, L_{IJ}$$

where now the flux is given by the normal derivative to the side IJ. Accordingly, we have along any boundary a total flux given by

$$F_{IJ} = k\, \frac{\mathrm{d}u}{\mathrm{d}n} + \boldsymbol{a} \cdot \boldsymbol{n}\, u \ .$$

A logical approximation for the normal derivative (which is horizontal in the figure) is given by the approximation

$$\frac{\mathrm{d}u}{\mathrm{d}n} \approx \frac{\tilde{u}^J - \tilde{u}^I}{h_{IJ}}$$

The convective part of the flux will again be taken as the *upwind* value, where we need again to consider the sign of $\boldsymbol{a} \cdot \boldsymbol{n}$ to decide on inflow and outflow parts. For triangle I we would use the value of $\tilde{u}^I$ if $\boldsymbol{a} \cdot \boldsymbol{n}$ is positive and that of $\tilde{u}^J$ if the product is negative. Once again, for this finite volume approach one gets a *full upwind* treatment.

The above steps are, in effect, the usual procedures for a cell centered finite volume approach and we observe that it has a certain commonality with the discontinuous Galerkin provided the upwinding term is always involved. If not the approximation is the more common one which is thus far distant from the finite difference process and still requires upwinding.

4.2 Node centered finite volumes

The whole procedure can be extended to node centered finite volumes but here more imaginative approximations may be required. To indicate why we again consider a one-dimensional application which again uses constant weighting over the element. However, now we admit C^0 continuous, linear variation for the solution variable u. A typical element is shown in Fig. 7. The balance equation is again given by Eq. (59), however, now it is possible to compute the flux variables directly from the linear interpolation between

the nodes. For example, the flux F_R may be expressed as

$$F_R = k\,\frac{\tilde{u}^{I+1} - \tilde{u}^I}{x^{I+1} - x^I} + \frac{1}{2}\,a\,\left(\tilde{u}^{I+1} + \tilde{u}^I\right) \tag{66}$$

and that for F_L by

$$F_L = k\,\frac{\tilde{u}^I - \tilde{u}^{I-1}}{x^I - x^{I-1}} + \frac{1}{2}\,a\,\left(\tilde{u}^I + \tilde{u}^{I-1}\right) \tag{67}$$

where we have again assumed that k and a are positive constants. The result for this approximation is clearly an average central difference approximation for *both* the convective and the diffusive fluxes. For equal size elements h we obtain the simple balance expression

$$\frac{k}{h}\left[u^{I-1} - 2\,u^I + u^{I+1}\right] + \frac{1}{2}\,a\,\left[u^{I+1} - u^{I-1}\right] = \bar{q}\,h \tag{68}$$

For problems with large convection content the solution will clearaly oscillate and some *upwind* treatment will be necessary.

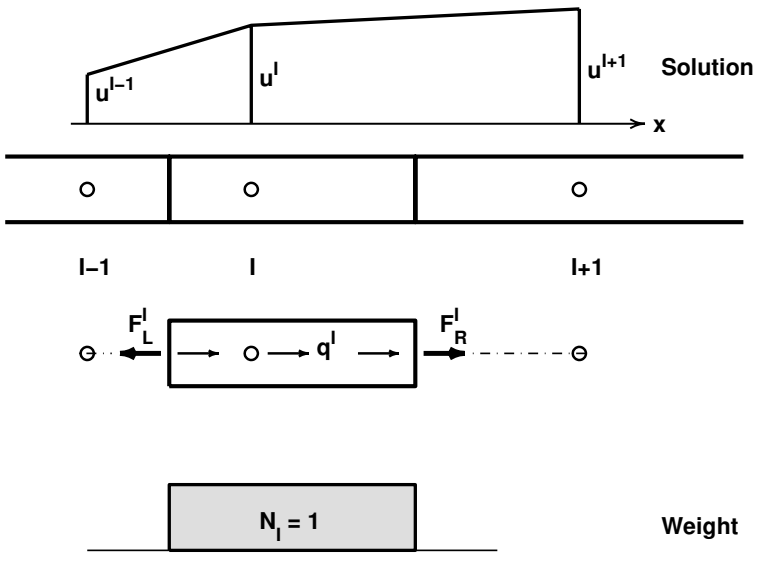

Figure 7: Node centered finite volume in one dimension

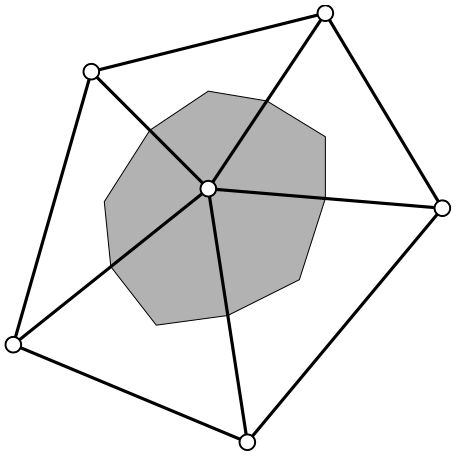

Figure 8: Node centered finite volume in two dimensions

Extension of the node centered approach to multiple dimensions is straight forward. For example, in Fig. 8 we show a set of triangles in which C^0 linear interpolation is made. The cell volume is constructed by connecting the midsides of the triangles with their centroids. The normal diffusive flux on each boundary segment may be computed by differentiating the linear shape functions (which gives a constant) and computing the constant normal to each boundary segment. Multiplying the result for the normal flux by the length and conductivity for each triangle gives the desired contribution to each segment. Similarly, the result for the convective flux may be easily computed; however, again we shall find that no *upwind* effect will be present and it will be necessary to devise some scheme to introduce one. The node centered approach described is clearly a mix between a discontinuous Galerkin approximation for the weight function and a continuous Galerkin approximation for the solution. One could devise an alternative scheme in which the types of interpolation are switched and in that case it would be possible to use an *upwind* approximation for the convective flux.

From the preceding it is seen that the finite volume approach if implemented with usual inflow and outflow type of boundary approximations to model correctly the convective terms can be classified in the same category as other elements of the discontinuous Galerkin type.

4.3 Node centered finite volume as a finite element method

A node centered finite volume method may be developed as a standard finite element type. That is, we may consider an element in which the force and stiffness terms are computed on a single element and assemble the result using a standard method.[3] Figure 9 shows a node centered finite volume for a typical node in a mesh of square elements. We have identified the parts of the contour as $ABCD$ and in the same figure for a single element shown how the contour parts can be computed as a finite element method. The final result will be the same whether we compute the complete contour for a given node at one time or if we compute the parts for the elements adjacent to the node separately and assemble to get the full node value.

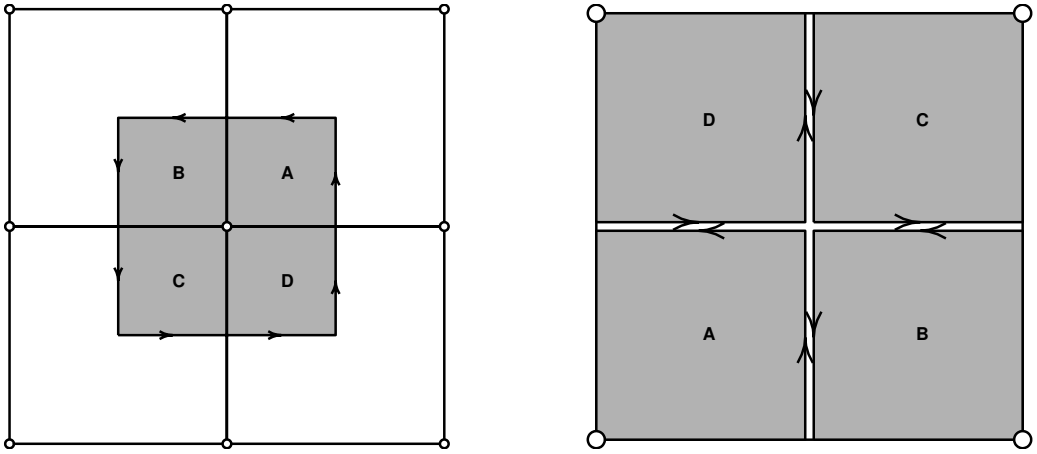

Figure 9: Node centered finite volume as a finite element

5 Numerical Examples

In this section we consider some example solutions for problems in one-dimension. The solutions are all computed in the context of special cases of the general linear diffusion-convection-reaction equation. However, to distinguish between some aspects of the various treatments we consider special

cases of this differential equation. The first set of examples consider a solution of problems in which convection effects are not included. Thus, the solution of the resulting diffusion-reaction equation belongs to a class of problems for which a variational theorem exists and continuous Galerkin approximations are an identical procedure for solving the problem. We thus shall be concerned with how well the discontinuous Galerkin method performs in comparison with standard finite element solutions in which C^0 approximating functions are used throughout. It is well known that such Galerkin (or variational) solutions are optimal in an energy sense.

In the second class of problems considered we consider the solution of problems which include convection effects only. Here, we explore how well the convective solution is given using the discontinuous Galerkin process. In the next set we add diffusion effects, first exploring cases where low convection is present, one in which standard Galerkin C^0 solutions perform quite well (e.g., ones in which low Peclet numbers are involved). We then look at the case in which diffusion effects are quite small (i.e., high Peclet numbers).

Finally we consider the case of a second problem which involves all effects. This is the Hemker problem described by Oden & Baumann[31] which has a turning point.

5.1 Diffusion-reaction example problem

We first consider the problem shown in Fig. 10 which is an example of a string under tension k and supported on an elastic Winkler foundation with support modulus c per length. The differential equation is given by

$$-k\,\frac{\mathrm{d}^2u}{\mathrm{d}x^2} + c\,u = q(x) \quad ; \quad -1 < x < 1$$

for which the properties $k = 1$ and $c = 9$ are used. A loading with intensity $q = 1$ is applied on $-0.1 < x < 0.1$ as shown in the figure.

The problem is solved by a simple mesh of 9 elements: 4 on each side of the loading and one for the loaded length. In Fig. 11 we show the result for a discontinuous Galerkin solution for which $p = 4$ and $\alpha = 1$ (Oden & Baumann unsymmetric method) is employed. This solution is nearly exact over the entire length.

For comparison we consider the solution to the problem using different order of approximation and both symmetric and unsymmetric treatment for the discontinuous Galerkin treatment of the diffusive flux. In Fig. 12 we

show the results for the solution using 9-quadratic order elements and both $\alpha \pm 1$. No τ stabilization is needed to get a stable solution (although the symmetric treatment resulted in changes in signs of diagonals in a direct solution without pivots; but little cancellation error in the magnitude of pivots).

In Fig. 13 we show a comparison between the 9-element $p = 2$ and $p = 3$ discontinuous Galerkin ($\alpha = 1$) and the same order classical C^0 finite element solution. For $p = 3$ we find a nearly converged solution for both methods and a plot cannot distinguish between results.

Finally, in graphical comparisons we present a pair of results for the case where linear interpolation is used in each element. First we present in Fig. 14 results for a 36-element mesh (where each element in the 9 element so-

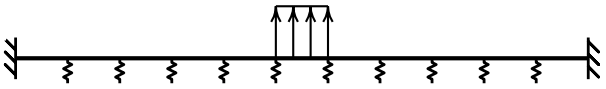

Figure 10: String on elastic foundation

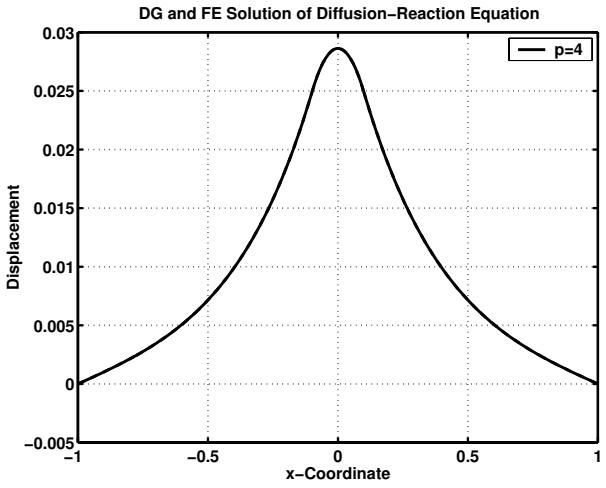

Figure 11: 9-Element DG solution for string on elastic foundation ($p = 4$ & $\alpha = 1$)

lution is divided into 4 equal subincrments). Here for the treatment shown it is clear that the discontinutity between elements is not completely eliminated. Indeed subsequent increases in the number of elements in the mesh by factors of 2 show very slow decay in the discontinuity. Oden & Baumann do not recommend use of linear approximation; however, treatment by the

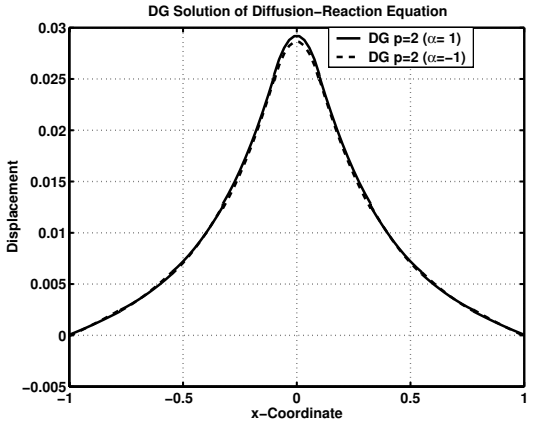

Figure 12: 9-element DG solutions for string on elastic foundation ($p = 2$ & $\alpha \pm 1$)

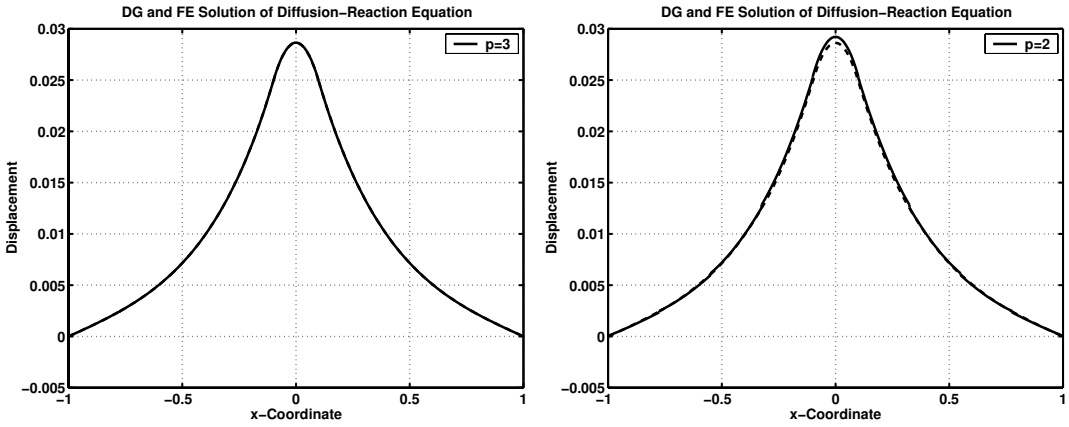

Figure 13: 9-element DG and FE solution for string on elastic foundation ($p =$, $p = 3$ & $\alpha = 1$)

symmetric discontinuous Galerkin treatment of flux ($\alpha = -1$) with non-zero τ gives some improvement to the solution as indicated in Fig. 15. In this application we also again consider the 9-element problem for which no stabilization is also presented in the figure. Results on 9-elements are nearly meaningless without the added τ stabilization term. In the computations shown $\tau \approx 5(k/h)$. Indeed it is easy to show that the discontinuous Galerkin solution approaches finite element solution when $\tau \to \infty$ as in that case continuity is restored by a *penalty process*. For comparison purposes we present in Fig. 16 the results for a standard finite element solution using linear approximations in each element. It is clear that superior results are attained using the standard finite element procedure.

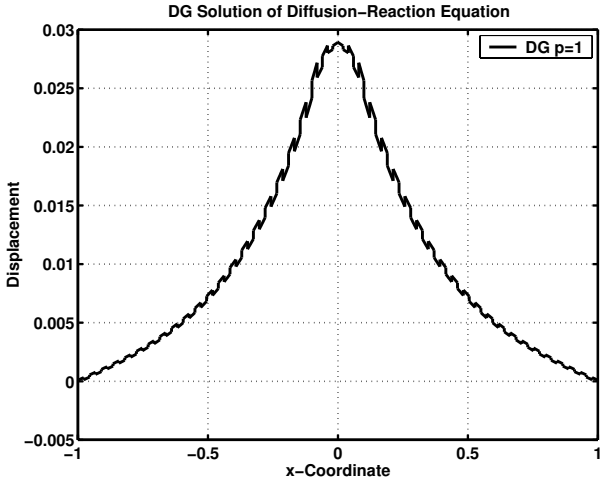

Figure 14: 36-element DG solution for string on elastic foundation ($p = 1$ & $\alpha = 1$)

In order to indicate the type of accuracy attained with the linear and quadratic order approximations we investigate the convergence of the displacement u at the center. The results are tabulated in Tables 1 and 2 for the linear and quadratic approximations, respectively. For the discontinuous Galerkin process no τ stabilization has been added. Indeed, as convergence occurred the discontinuity between elements did become smaller and except for linear case was negligible long before the final mesh was attained.

Several conclusions can be reached from the results presented from this problem. First, the symmetric treatment of the flux leads to better accuracy.

This might have been expected as the basis for the solution is rooted in a full variational theorem (with added stabilized terms). Secondly, as also could be anticipated, the standard finite element solution process for this

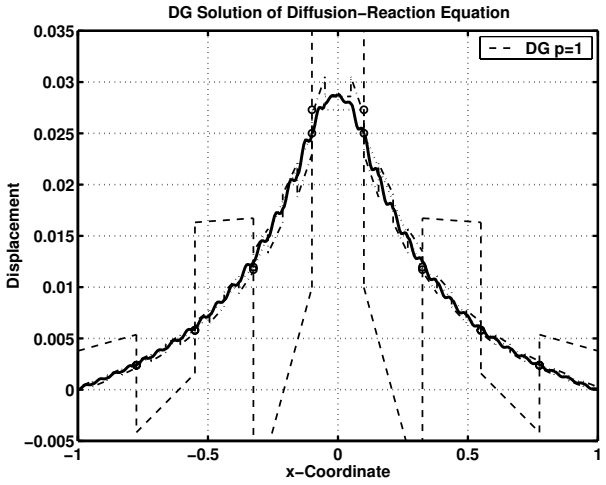

Figure 15: 9 and 36-element DG solution for string on elastic foundation $[p = 1,\ \alpha = -1$ and $\tau \approx 5(k/h)]$

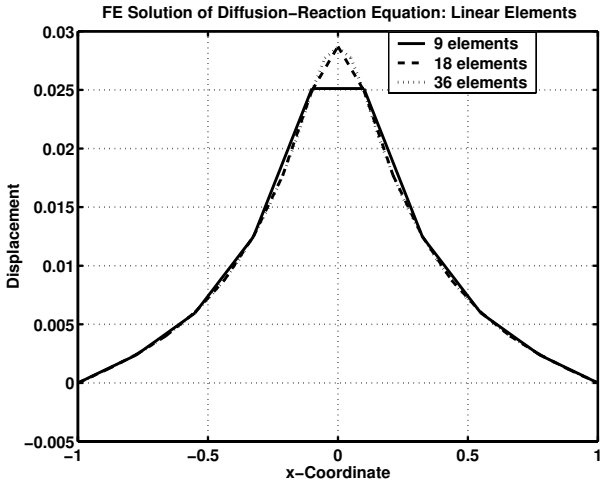

Figure 16: 9 to 36 element FE solution for string on elastic foundation ($p = 1$)

N_{elm}	FE	DG($\alpha = -1$)	DG($\alpha = 1$)
9	0.025117	-0.121829	0.059726
18	0.028710	0.028709	0.035672
36	0.028650	0.028634	0.028575
72	0.028636	0.028632	0.028617
144	0.028632	0.028631	0.028627

Table 1: String on elastic foundation: Solution by FE and DG for $p = 1$

N_{elm}	FE	DG($\alpha = -1$)	DG($\alpha = 1$)
9	0.028634	0.028667	0.029224
18	0.028630	0.028630	0.028808
36	0.028631	0.028631	0.028684
72	0.028631	0.028631	0.028645
144	0.028631	0.028631	0.028634

Table 2: String on elastic foundation: Solution by FE and DG for $p = 2$

class of problems is far more efficient and far more accurate for a given number of elements. When compared on a basis of the total number equations involved the comparison is even more favorable to the standard finite element treatment.

5.2 Pure convection example

As an example of pure convection we consider the problem

$$\frac{\mathrm{d}u}{\mathrm{d}x} = q(x) \quad \text{for } u(0) = 0$$

with the loading specified as shown in Fig. 17. A solution obtained using the discontinuous Galerkin method for approximations of order $p = 1$ and 2 is also shown in Fig. 17. Since the loading varies linearly with position, the exact solution is composed of piecewise polynomials of degree two. Hence, the discontinuous Galerkin solution with $p = 2$ yields the exact solution and use of higher order approximations is unnecessary. The linear approximation, on the other hand, leads to discontinuitites at the element interfaces. Use of

smaller elements would reduce the size of the jumps. Finally, for this problem no diffusion effects are present and, thus, there is no effect on choice of α.

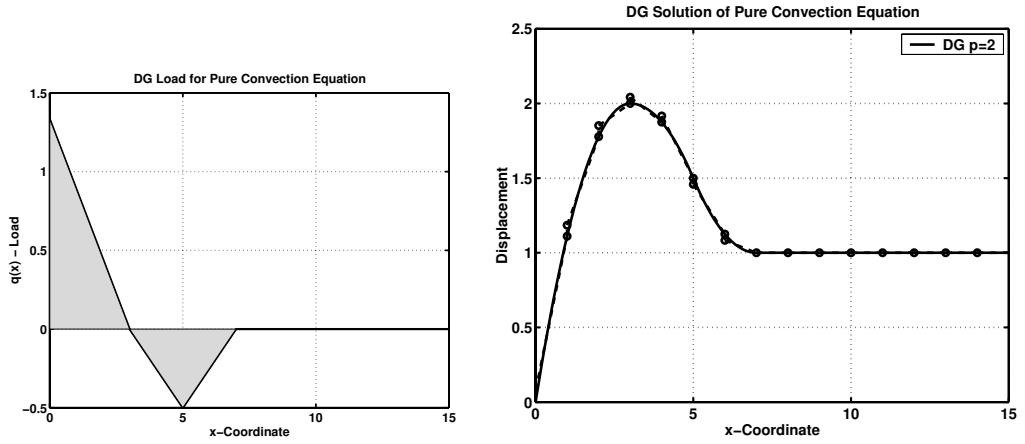

Figure 17: Loading and DG solution for pure convection example for $p = 1$ and $p = 2$

5.3 Convection-diffusion example

The first problem selected for this class is a classical convection-diffusion problem given by the differential equation

$$\frac{\mathrm{d}^2 u}{\mathrm{d}x^2} + b\,\frac{\mathrm{d}u}{\mathrm{d}x} = 0 \quad ; \quad 0 < x < L$$

with Dirichlet boundary conditions $u(0) = 1$ and $u(L) = 0$.

We first compute a converged solution for the case $b = 20$ and $L = 10$ using $p = 3$ and $\alpha = 1$ (shown in Fig. 18). For comparison purposes we also compute standard finite element solutions for meshes of 10 elements with linear, quadratic, cubic and quartic C^0 interpolations. Results for these are given in Fig. 19. As no upwind modifications are included we note that the solutions are oscilatory. Indeed these could be improved by introducing an upwind strategy. A solution using *optimal* upwind treatment[7] for linear elements is first computed. Results are also computed for the quadratic order element ($p = 2$) with upwind treatment taken as the optimal divided by two

(higher order elements would divide by p). Results for this comparison are presented in Fig. 19.

For the same example we compute solutions using the discontinuous Galerkin procedures described above with $\alpha = 1$ and polynomial orders ranging from 1 to 4. Results are given in Fig. 21.

As a second example we consider the problem above for various values of the parameter b. The mesh is selected with 9 elements, each with length h of 2. In this case, for linear elements ($p = 1$), the value of b is precisely the element Peclet number given by[7]

$$Pe = \frac{b\,h}{2} \ .$$

The problem is first solved by standard finite elements (without upwind treatment) using linear elements of equal length and values of the Peclet number of 0, 1, 2.5, and ∞. The results are shown in the left diagram in Fig. 22. The problem is repeated using optimal upwinding and, as expected, produces the exact values at nodes for all Peclet numbers but with linear interpolations between the nodes.

This problem is next analyzed by the Discontinuous Galerkin method with $p = 1$ and again 9-elements. In addition the problem is analyzed for a cubic order approximation ($p = 3$) but using only 3-elements. Results for these two analyses are presented in Fig. 23. The results obtained are not

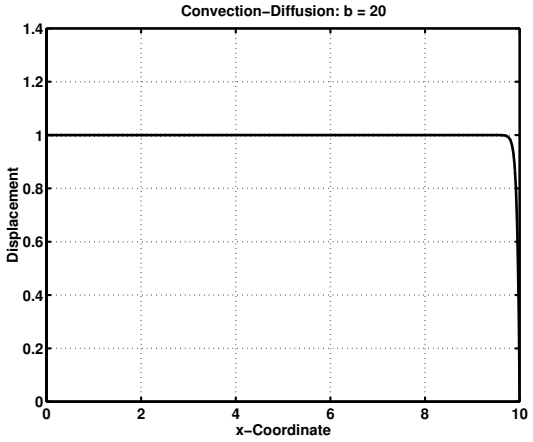

Figure 18: DG solution for convection-diffusion example for $p = 3$

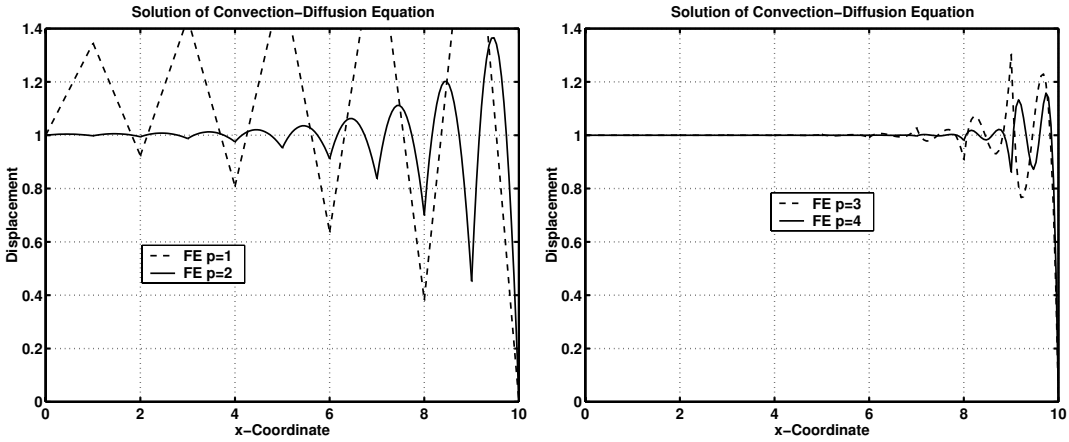

Figure 19: 10-element FE solution for convection-diffusion example using $p = 1$ to $p = 4$

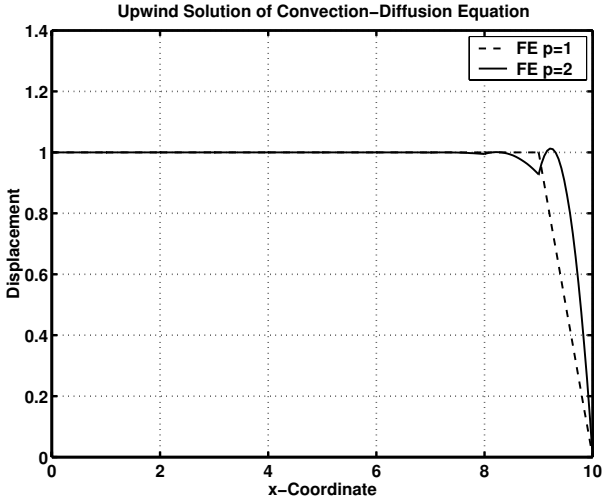

Figure 20: 10-element upwind FE solution for convection-diffusion example using $p = 1$ and $p = 2$

accurate for any of the non-zero values of the Peclet number. Indeed, cubic results do not lead to an improved estimate of the response. To examine

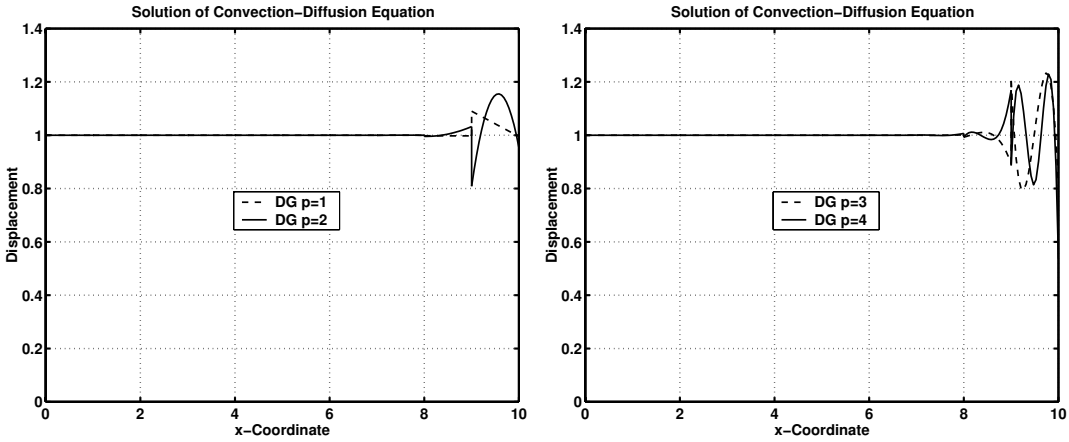

Figure 21: 10-element DG solution for convection-diffusion example using $p = 1$ to $p = 4$

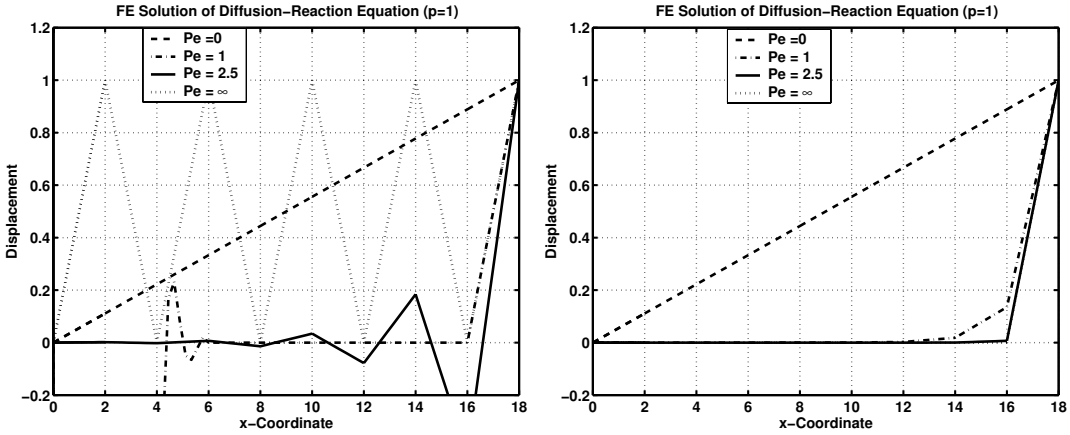

Figure 22: 9-element FE solution for convection-diffusion with $Pe = 0, 1, 2.5, \infty$ $(p = 1)$

the behavior with a mesh in which the element size varies by a large ratio between the left and right boundaries we consider a 9-element non-uniform $p = 1$ and a 3-element $p = 3$ example. The mesh was generated by using

equal increments of ξ for the quadratic isoparametric interpolation given by

$$x = (1 - \xi^2)\, 14 + \frac{1}{2}\,(\xi + \xi^2)\, 18 \ .$$

The results of the analysis are displayed Fig. 24 and are found to now represent quite accurately (though note exactly for all Peclet numbers) the results. It is remarkable the amount by which the discontinuities are reduced by using a highly graded mesh.

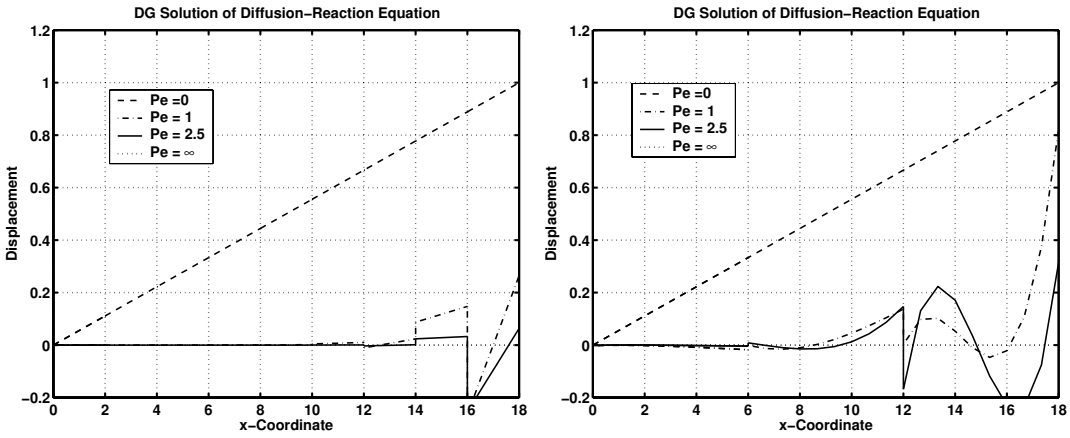

Figure 23: 9-element $p = 1$ and 3-element $p = 3$ DG solution for convection-diffusion with $Pe = 0, 1, 2.5, \infty$ $(p = 1)$

5.4 Hemker problem example

One example presented by Oden $ Baumann[31] is the Hemker problem. The differential equation for this problem is given by

$$-k\,\frac{\mathrm{d}^2 u}{\mathrm{d}x^2} - x\,\frac{\mathrm{d}u}{\mathrm{d}x} = f \quad ; \quad -1 < x < 1$$

where the load function is given by $f = k\pi^2 \cos(\pi x) + \pi x \sin(\pi x)$. The boundary conditions for the problem are both Dirichlet with $u(-1) = 2$ and $u(1) = 0$. The exact solution for the problem as reported by Oden & Baumann is

$$u(x) = cos(\pi x) + \mathrm{erf}(x/\sqrt{(}2k))/\mathrm{erf}(1/\sqrt{(}2k))$$

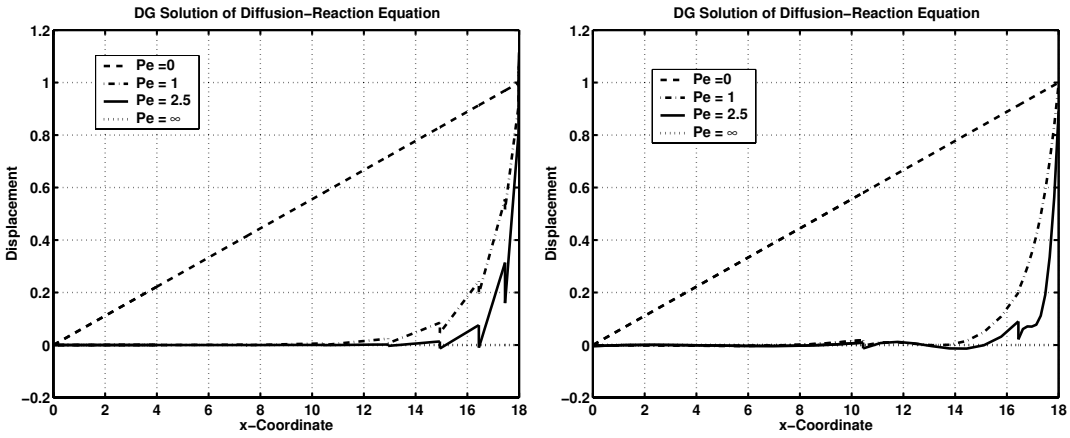

Figure 24: Non-uniform 9-element $p = 1$ and 3-element $p = 3$ DG solution for convection-diffusion with $Pe = 0, 1, 2.5, \infty$ $(p = 1)$

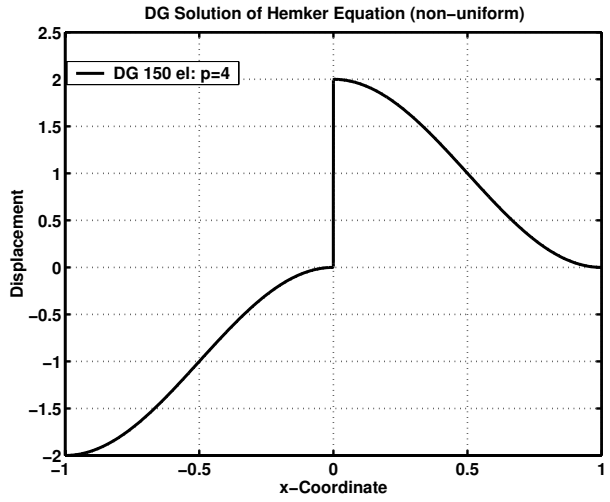

Figure 25: DG solution of Hemker example for $p = 4$ and 150 elements

In the reference cited it is assumed that $k = 10^{-10}$ and thus, the error function erf produces a near vertical jump of magnitude 2 at $x = 0$ with the remainder

of the solution given by the cosine term.[2]. A quite accurate representation of this solution is given in Fig. 25 where a discontinuous Galerkin solution using quartic order interpolation in elements and 150 elements is presented.

We also compute standard finite element approximations for the solution using linear and quadratic order interpolations with and without upwinding as described above. For the analysis we use 16-linear elements and 8-quadratic elements evenly distributed along the length. Results are given in Fig. 26. To indicate what happens as the number of elements is increased we also present a solution using 200 linear elements (Fig. 27) which include upwinding (without upwinding the solution oscillates significantly near the discontinuity). Optimal upwinding for the constant case is used at each quadrature point and this is seen to be quite effective with only a small Gibbs type overshoot near the jump. The solution is, however, somewhat damped from the pure cosine response (which should be 0 before the jump and 2 after). Thus, use of upwinding has a negative impact on the solution. Comparison between Fig. 26 and 27 does indicate a reduction in this damping effect is occuring and, thus, convergence will eventually occur.

The analysis is now repeated using discontinuous Galerkin procedures with $p = 1$ to $p = 4$ and a uniform mesh of 16 elements for the linear case and 8-elements for all the others. This is not a fair comparison with the finite element solutions on an effort basis as the discontinuous approximation involves one additional unknown for each interface or boundary point, however, on a polynomial order basis the comparison is useful.

We now consider the solution to the same problem using a discontinuous Galerkin procedure with $\alpha = 1$ (the Oden & Baumann treatment). Meshes of 16 elements for $p = 1$ and 8 elmeents for higher order p are used with results shown in Fig. 28. It is evident that the discontinuous Galerkin method is nearly ideal for this problem and reproduces faithfully the cosine like solution with the proper jump given. Indeed, one may inquire to what degree one could perturb the location of the element end from the jump location without affecting the quality of the solution. Such an experiment has been conducted for a misplacement of $e = -0.01$ units and results are shown in Fig. 29. Surprisingly, there is little polution of the solution beyond the one element affected by the perturbation. Repeating the experiment for the finite element solution with linear elements has a similar effect, however,

[2]The solution is continuous at $x = 0$ the overall effect of the error function is restricted to a very narrow band around this point

the quality of the original solution is not good so neither are those for the perturbed solution. The results are shown in Fig. 30

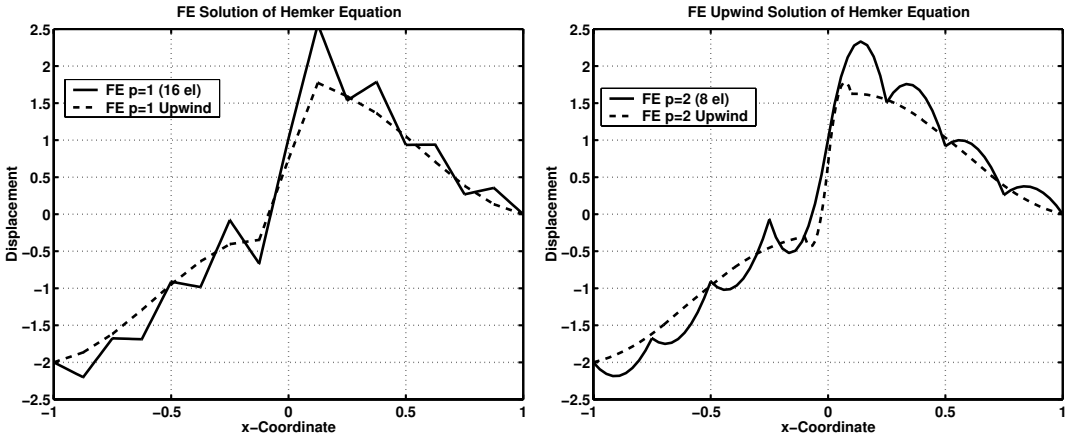

Figure 26: FE solution for Hemker example using $p = 1$ and $p = 2$

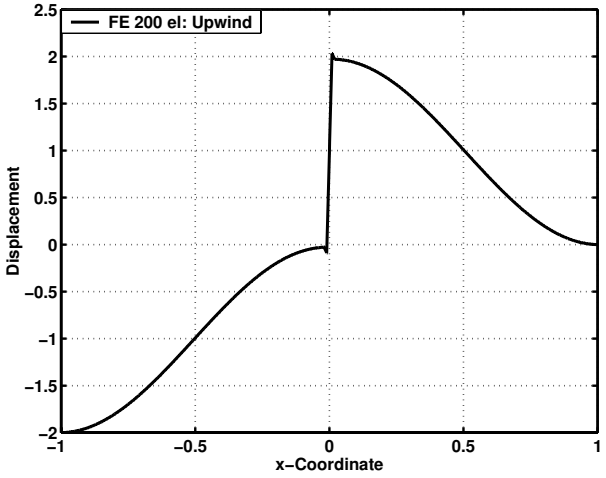

Figure 27: FE solution for Hemker example using $p = 1$ and 200 linear elements

6 Closure

In this report we have considered solution by discontinuous Galerkin methods of problems involving diffusion, convection and reaction. The general form of

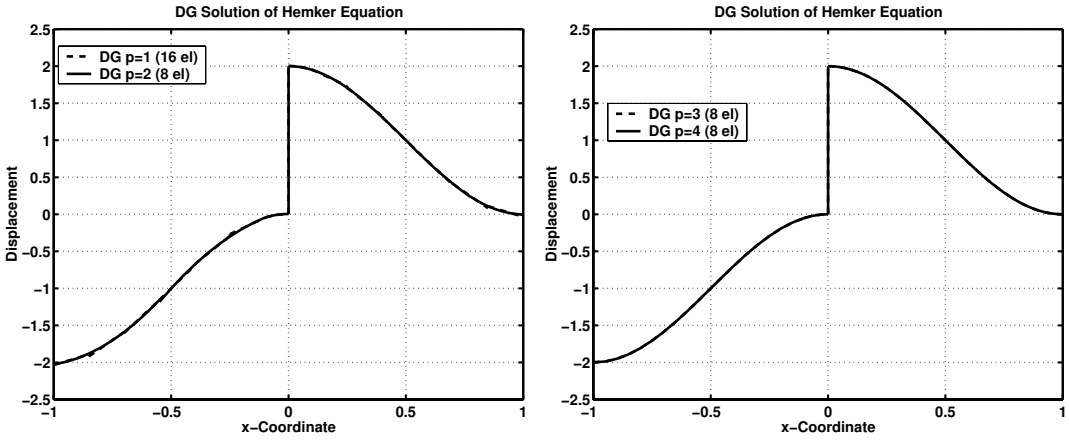

Figure 28: DG solution for Hemker example using $p = 1$ to $p = 4$

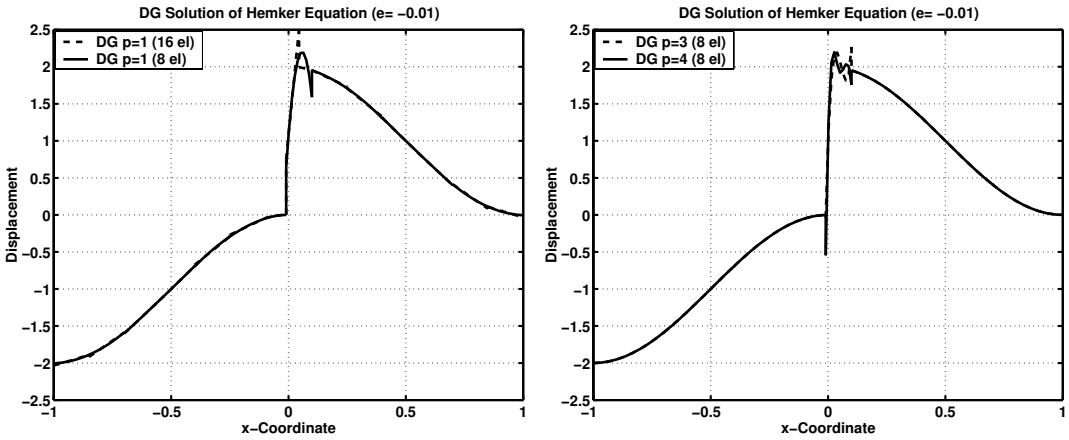

Figure 29: DG solution of Hemker example for $e = -0.01$ using $p = 1$ to $p = 4$

the discontinuous Galerkin procedure for a scalar equation has been presented along with alternatives for treating the convective and diffusive fluxes. A comparison with the finite volume method has also been discussed.

Numerical examples are presented for one dimensional applications. Here we have compared the symmetric treatment of the diffusive flux with an unsymmetric form proposed by Oden & Baumann. Our assessment shows that the symmetric treatment is more accurate than the unsymmetric one, however, this is balanced by additional stability obtained by the unsymmetric treatment. Differences in accuracy are small once third order approximations are used ($p = 3$) in each element but are noticable for second order treatment. For the diffusion-reaction problem the best accuracy results from standard finite element treatment using C^0 approximation throughout.

Analysis of convective terms is included using an *upwind* treatment at the interfaces. This is shown to be very effective in correctly transmitting the convective flux throughout the domain, but is sensitive to disturbances introduced. There is need for additional study on effective means to introduce some upwind treatment within the individual domains (elements) in a manner which does not destroy balance properties. The Hemker problem shows that traditional treatment by upwind finite element methods is not very effective. Significant error is obtained around the turning point and only very slowly

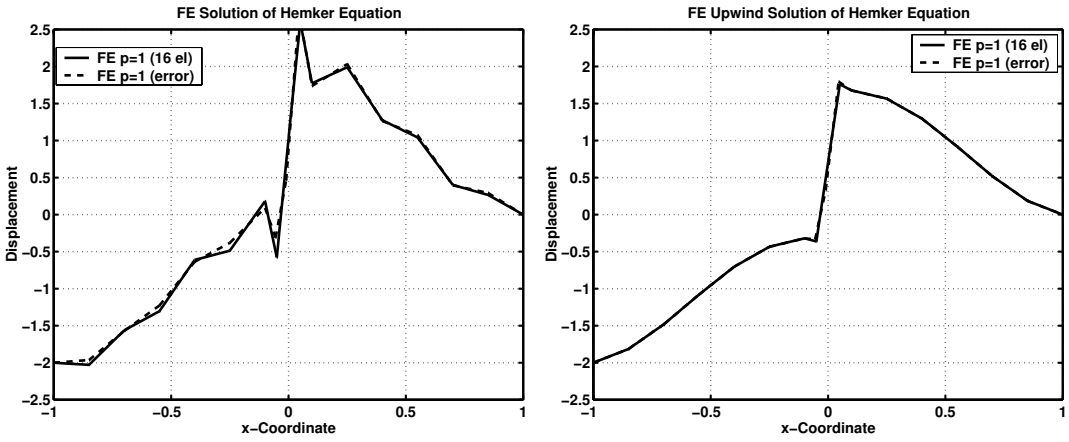

Figure 30: FE solution of Hemker example for $e = -0.01$ using $p = 1$ and 16 elements

is eliminated by increasing the number of elements. Moreover, use of higher order approximations is not easily treated by traditional means.

An important observation from the simple examples treated in this report that any error in placement of the point where sharp gradients in solution occur (e.g., jumps) does not polute the solution at distant points from the misplacement. This is in contrast to conventional finite element errors which do often polute the entire solution. Whether this observation is valid in other cases needs additional study.

Another advantage of the discontinuous Galerkin method is the structure of the resulting mass type matrix. Here the matrix is of block diagonal structure and can, as done by Karniadakis *et al.*, be put in diagonal form by suitable construction of orthogonal polynomial approximations. Thus, p-order approximations may be constructed with fully diagonal matrices. Of course this is only of importance in explicit time integration problems but perhaps can be explointed further in iterative approaches.

Finally, we again remark that the advantages of the discontinuous Galerkin method do not come without some additional costs. First, the number of variables is increased by the introduction of discontinuous approximations. Secondly, the overall structure of the implementation is complicated over that of standard finite element procedures. This is evident in the need to treat the interfaces where information from contiguous elements is needed to construct the flux terms. Also, we have found that some problems have indefinite form even using the Oden & Baumann treatment. Some preliminary solutions using other values of the α than plus or minus unity can be effective. Thus, another avenue for further study is available.

References

[1] B.G. Galerkin. Rods and Plates. Series occuring in various questions conderning the elastic equilibrium of rods and plates (Sterzhni i plastiny. Ryady v nekotorykh voprosakh uprogogo ravnovesiya sterzhnei i plastin.). *Engineers Bulletin (Vestnik inzhenerov)*, 19:897–908, 1915.

[2] S. Crandall. *Engineering Analysis*. McGraw-Hill, New York, 1956.

[3] O.C. Zienkiewicz and R.L. Taylor. *The Finite Element Method: The Basis*, volume 1. Butterworth-Heinemann, Oxford, 5th edition, 2000.

[4] I.G. Bubnov. Report on the works of Prof. Timoshenko which were awarded the Zhuranskii prize. Symposium of the Institute of Communication Engineers *(Sborn. inta inzh. putei soobshch.)*. Technical Report No. 81, All Union Special Planning Office (SPB), 1913.

[5] G.I. Petrov. Application of the method of Galerkin to a problem involving the stationary flow of a viscous fluid. *Prikl. matem. i mekh.*, 4:3, 1940.

[6] S.G. Mikhlin. *Variational Methods in Mathematical Physics*. Macmillan, New York, 1964.

[7] O.C. Zienkiewicz and R.L. Taylor. *The Finite Element Method: Fluid Mechanics*, volume 3. Butterworth-Heinemann, Oxford, 5th edition, 2000.

[8] J. Donea. A Taylor-Galerkin method for convective transport problems. *International Journal for Numerical Methods in Engineering*, 20:101–119, 1984.

[9] R. Löhner, K. Morgan, and O.C. Zienkiewicz. The solution of non-linear hyperbolic equation systems by the finite element method. *International Journal for Numerical Methods in Fluids*, 4:1043–1063, 1984.

[10] T.J.R. Hughes, L.P. Franca, and G.M. Hulbert. A new finite element formulation for computational fluid dynamics: VIII. The Galerkin/least-squares method for advective-diffusive equations. *Computer Methods in Applied Mechanics and Engineering*, 73:173–189, 1989.

[11] M. Delfour and F. Trochu. Discontinuous Galerkin methods for the approximation of optimal control problems governed by hereditary differential systems. In A. Ruberti, editor, *Distributed Parameter Systems: Modelling and Identification*, pages 256–271. Springer-Verlag, 1978.

[12] C. Johnson and J. Pitkäranta. An analysis of the discontinuous galerkin method for a scalar hyperbolic equation. *Math. Comp.*, 46:1–26, 1986.

[13] C. Johnson. *Numerical Solutions of Partial Differential Equations by the Finite Element Method*. Cambridge University Press, Cambridge, 1987.

[14] T.H.H. Pian. Derivation of element stiffness matrices by assumed stress distribution. *Journal of AIAA*, 2:1332–1336, 1964.

[15] T.H.H. Pian and P. Tong. Basis of finite element methods for solid continua. *International Journal for Numerical Methods in Engineering*, 1:3–28, 1969.

[16] F. Kikuchi and Y. Ando. A new variational functional for the finite element method and its application to plate and shell problems. *Nuclear Engineering and Design*, 21(1):95–113, 1972.

[17] G.P. Bazeley, Y.K. Cheung, B.M. Irons, and O.C. Zienkiewicz. Triangular elements in bending – conforming and non-conforming solutions. In *Proc. 1st Conf. Matrix Methods in Structural Mechanics*, volume AFFDL-TR-66-80, pages 547–576, Wright Patterson Air force Base, Ohio, October 1966.

[18] J.A. Nitsche. Über ein Variationsprinzip zur Lösung Dirichlet-Problemen bei Verwendung von Teilräumen, die keinen Randbedingungen uneworfen sind. *Abh. Math. Sem. Univ. Hamburg*, 36:9–15, 1971.

[19] W.H. Reed and T.R. Hill. Triangular mesh methods for the neutron transport equation. Technical Report Technical Report LA-UR-73-479, Los Alamos Scientific Laboratory, 1973.

[20] P. Lesaint and P.-A. Raviart. On a finite element method for solving the neutron transport equation. In C. de Boor, editor, *Mathematical Aspects of Finite Elements in Partial Differential Equations*. Academic Press, New York, 1974.

[21] G.E. Karniadakis and S.J. Sherwin. *Spectral/hp Element Methods in CFD*. Oxford University Press, Oxford, 1999.

[22] T.C. Warburton, I. Lomtev, R.M. Kirby, and G.E. Karniadakis. A discontinuous Galerkin method for the Navier-Stokes equations in hybrid grids. In M. Hafez and J.C. Heinrich, editors, *10th. International Conference on Finite Element Methods in Fluids*, Tucson, Arizona, 1998.

[23] T.C. Warburton and G.E. Karniadakis. A discontinuous Galerkin method for the viscous MHD equations. *J. Comput. Physics*, 152:1–34, 1999.

[24] B. Cockburn. An introduction to the discontinuous Galerkin method for convection-dominated problems. In A. Quarteroni, editor, *Advanced numerical approximation of nonlinear hyperbolic equations*, volume 1697 of Lecture Notes in Mathematics, pages 151–268. Springer-Verlag, 1998.

[25] B. Cockburn. Discontinuous Galerkin methods for convection-dominated problems. In T. Barth and H. Deconink, editors, *High-Order Methods for Computational Physics*, volume 9 of Lecture Notes in Computational Science and Engineering, pages 69–224. Springer-Verlag, 1999.

[26] C.E. Baumann. *An hp-Adaptive Discontinuous Finite Element Method for Computational Fluid Dynamics*. Ph.d dissertation, The University of Texas, Austin, Texas, 1997.

[27] J.T. Oden, I. Babuška, and C.E. Baumann. A discontinuous *hp* finite element method for diffusion problems. *J. Comp. Physics*, 146(2):491–519, 1998.

[28] J.T. Oden and C.E. Baumann. A discontinuous *hp* finite element method for convection-diffusion problems. *Computer Methods in Applied Mechanics and Engineering*, 175(3-4):311–341, 1999.

[29] C.E. Baumann and J.T. Oden. A discontinuous hp finite element method for convection-diffusion problems. *Computer Methods in Applied Mechanics and Engineering*, 175:311–341, 1999.

[30] C.E. Baumann and J.T. Oden. A discontinuous hp finite element method for the Euler and Navier-Stokes problems. *International Journal for Numerical Methods in Engineering*, page in press, 1999.

[31] J.T. Oden and C.E. Baumann. A conservative DGM for convection-diffusion and Navier-Stokes problems. In *Discontinous Galerkin Methods: Theory, Computation and Applications*, pages 179–194, Berlin, 2000. Springer-Verlag.

[32] T.J.R. Hughes, G. Engel, L. Mazzei, and M.G. Larson. A comparison of discontinuous and continuous Galerkin methods based on error estimates, conservation, robustness and efficiency. In *Discontinous Galerkin Methods: Theory, Computation and Applications*, pages 135–146, Berlin, 2000. Springer-Verlag.

[33] B. Cockburn, G.E. Karniadakis, and Chi-Wang Shu. *Discontinous Galerkin Methods: Theory, Computation and Applications*. Springer-Verlag, Berlin, 2000.

[34] C. Johnson, U. Nävert, and J. Pitkäranta. Finite element methods for linear hyperbolic problems. *Computer Methods in Applied Mechanics and Engineering*, 45:285–312, 1984.